SILVIA TOSSUT
Università Vita-Salute san Raffaele
silvia.tossut@gmail.com

ON ACTING *BECAUSE OF* A JOINT COMMITMENT

abstract

I focus on the motivational role that Gilbert attributes to joint commitment. Using Bacharach's game theoretical analysis of non-summative group agency, I point out that Gilbert's account of social actions conceptually requires the obliteration of individual preferences and individual rationality. Then, I investigate whether acting because of a joint commitment is rational in some sense, focusing primarily on the phenomenon of asking and giving the permission to defect (fair defection) when a joint commitment is in place. I show that the obliteration of individual preferences prevents the possibility of rational fair defection. Finally, I analyze Gilbert's recent suggestion concerning the introduction of personal preferences in her account, and I show that such introduction cannot solve the problem with the (ir)rationality of fair defection, and that personal preferences can hardly be consistently included in Gilbert's account.

keywords

joint commitment, game theory, Bacharach, defection

Phenomenology and Mind, n. 9 - 2015, pp. 38-45
DOI: 10.13128/Phe_Mi-18150
Web: www.fupress.net/index.php/pam

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

1. Gilbert's classical account of joint commitment

In this paper I assume the general validity of Margaret Gilbert's account, and I focus on a specific problem raising in this framework. The question that I will try to answer is: When I am part of a plural subject, what motivates my actions? What should I *rationally* do?
According to Gilbert, when a joint commitment is in place the people involved must act having this joint commitment as their motivation. Her thesis is that subjection to a joint commitment compels me from a *rational* point of view to act in a way consistent with its content: neither me nor my partners in the activity can remove the joint commitment at will, so I have sufficient reasons to undertake actions consistent with the collective goal expressed by the joint commitment. In Gilbert's words, once people are jointly committed to pursue a goal as a body, the group's goal becomes the unique motivational source for the actions of each of the parties (Gilbert 2006, p. 123), and thus, "a member's action may often be explained without any reference to his or her own personal goals, values, or principles of action" (Gilbert 1996, p. 268). As a consequence, the description of social actions in terms of joint commitment has the advantage of providing a *unified explanation* of the individual contributions to the action, even in the complete absence of reference to personal intentions or desires (Gilbert 2000, pp. 14-36). In other words, when a joint commitment is in place people involved act *because of* their joint commitment. A rational agent acting as a part of a plural subject must behave consistently with the intention to achieve the plural goal, and her behavior must be motivated by the joint commitment.
It is worth noting that the importance of joint commitment consists precisely in its functioning as a unified motivational source, since this unity confers stability to the cooperative outcome. Even though individual rationality can succeed in prescribing cooperation when specific circumstances occur, the existence of a joint commitment allows me to rely on the *assumption* that the others will cooperate. On the contrary, in the individualistic frame I must always consider the possibility that my partner might abandon cooperation.
The explanation of this kind of stability is *psychological*: as List and Pettit (2011, p. 193) put it, when people act in their capacity as group members, they experience a change in the perceived subject of intention and action. The reason is that the pronoun "we" has the power to provoke an immediate psychological response in the subject, inducing one's self-identification as a member of the group. The members of a group are thus the indivisible parts of a single center of agency.

2. A game-theoretical explanation of joint commitment

The departure from the singularist approach is also embraced (from a different perspective) by Michael Bacharach (2006), who presents a game theoretical analysis of cooperation quite sympathetic with Gilbert's plural subject theory. Bacharach's analysis is particularly useful to analyze the rationality of acting *because of* a joint commitment.

Bacharach's starting assumptions are (i) that if a group is an agent, then it has its own payoff function (distinct from the payoff functions of its members), and (ii) that, if rational, the group agent will aim at the maximization of *its* expected utility. According to Bacharach – and consistently with Gilbert's holistic perspective – the group's preferences cannot be reduced to the members' preferences, nor can the reasons motivating the group's choice be reduced to the members' reasons. Roughly, cooperative outcomes are explained by the individual agents propensity to *group identify*.

Group identification is a psychological response, primed by some objective features of the game (i.e., the immediate perception of a common interest and of the necessity of the others' contributions).[1] Group identification involves a shift of frame, a change in the agent's conception of herself: a player who group-identifies conceives herself as part of a unified agent (Bacharach 2006, p. 70). The crucial effects of group identification are payoff transformation, cooperation, and agency transformation.

Payoff transformation corresponds to the unification of the motivational sources that is prompted by group identification. Since all the actions of the players are motivated by the group's goal, group identification induces also a *cooperative behavior* within the group, (Bacharach 2006, p. 79). The mechanism that conveys these effects, and which is at the core of Bacharach's account, is the *agency transformation* prompted by group identification. Agency transformation consists in the players' adoption of a novel perspective with respect to the game. In summary, group identification causes the *disappearance* of personal payoffs, and a re-description of the game. (Compare this with Gilbert's idea that being part of a plural subject causes a change in the perceived subject – from the "I" to the "we").

To explain this point, let me consider the game in *figure 1*, which is a "Prisoner's Dilemma" (PD). In PD, each player chooses among "cooperate" (C) or "defect" (D). Given the structure of the game, each player prefers the outcome in which she defects and the other cooperates

		P1	
		C	D
P2	C	2,2	-3,3
	D	3,-3	1,1

Figure 1

	C	D
C	4	0
D	0	2

Figure 2

1 Bacharach takes identification with the group to be basically a non-rational framework effect, so it is not clear where the commitment is supposed to come from. This seems to be an important point to explain the integration of Gilbert's notion of joint commitment within Bacharach's account. I think that this feature is consistent with Gilbert's account to the extent that it is not always clear *why* individuals do enter a joint commitment. In what follows I offer an analysis of the motivations leading to the formation of a joint commitment, at least for those cases in which the framework effect is weak enough to leave space for a voluntary decision. I thank an anonymous referee for suggesting me to clarify this point.

("free-riding" situation); conversely, each player aims at avoiding the situation in which she cooperates and the other defects. The application of individual rationality leads to the solution (D,D). Yet, despite this prescription, both players might obtain higher payoffs if both cooperate.

Bacharach's analysis of cooperation lies on the assumption that group identification causes the players' understanding the PD as the game depicted in *figure 2*: the payoffs of this matrix are the payoffs of a *group agent*, which are achievable through the implementation of a certain profile of action by the original players.[2]

Note that the disappearance of personal payoffs and the new matrix of the game are not merely due to a change in the preferences of the individual agents: it is not simply that a group-identifier wishes spontaneously to promote the collective goal, rather she "thinks of her agential self [...] as a component part of [the group's] agency" (Bacharach 2006, p. 136). Agency transformation entails a change from the payoffs that govern choices for one unit of agency (the individual), in the payoffs that govern choices for another unit of agency (the group). Group identification amounts to feeling a sense of collectivity, which prompts "team-reasoning". Team-reasoning is possible only if the players are in a frame in which first-person plural concepts are activated (Bacharach 2006, p. 135 and p. 141): "if a group-identifier thinks of herself as part of a 'we' [...] it is only for us that she can intelligibly deliberate" (Bacharach 2006, p. 145). The peculiar form of reasoning entailed by group identification, also, supports the holistic interpretation of the "we", which cannot be reduced to the sum of individual motivations and actions: team-reasoning eliminates any reference to the individual, and allows the players to adopt a genuinely collective perspective.

The profile of action (C,C) in *figure 2* – i.e., from the point of view of the collective agent – is *insensitive* to the distribution of gain among the participants.[3] As Gilbert correctly points out, the obliteration of individual preferences due to the adoption of the plural subject perspective is a phenomenon that we can observe more clearly in particular kinds of groups (e.g., in marriage, in which people are likely to arrive to what she calls the "fusion of egos"). But the question is: besides being descriptively accurate, is the obliteration of individual preferences also *rational*?

In particular, Bacharach would say that the cooperative outcome (C,C) in PD *is* rational if we consider the collective agent as the subject of rationality, since (C,C) is the profile of action that maximizes its utility. Group identification transforms the original PD in the game in *figure 2*, which is a game played by the collective agent, and which has a unique "collective solution", the profile of action (C,C).

The peculiarity of the collective solution of the PD is that it is *inconsistent* with the solution prescribed by individual rationality. According to Bacharach, this inconsistency influences the final outcome, to the extent that the players' preferences toward a non-cooperative outcome lower the probability that they will group identify.

Nonetheless, in competitive frameworks, I can still rely on the other's doing her part in the realization of the collective solution. Indeed, Bacharach claims that since group identification is an involuntary psychological response, activated by *objective features* of the situation, in framing the situation as a problem "for us", an individual also gains some sense of *how likely it is* that another

2 The discussion is open on the exact mathematical representation of the collective payoff. For example, it might not be a sum, but rather an average of the individual payoffs. What I am claiming here is just that, whatsoever form this measure may taken, it will always be insensitive to individual preferences.

3 A complete account would require the analysis of asymmetric PD games, since such games entail a number of different considerations (e.g., one of the player could justify occasional defection on the basis of egalitarian considerations). Alas, such a complete analysis goes beyond the limits of this paper.

agent, facing the same situation, will we-frame it. Arguably, the more competitive is a game, the less intuitive – and the less probable – will be group identification (Bacharach 2006, p. 75).
The obliteration of the individual payoffs entails that the players might be required to act in contrast with the prescriptions of individual rationality. As in the PD, the adoption of the collective perspective results in an efficient solution. An objector might observe that, though the solution (C,C) is *prima facie* efficient for each player since it allows avoiding the worst profile (D,D), it is easy to see that each player can still maximize her expected utility: if I can *assume* that you will do your part in the collective solution, and I am rational, I must defect to obtain the profile (D,C). The answer to this objection is that "collective efficiency" is insensitive to the players' individual payoffs – consistently with Gilbert's non-summative approach to collective notions. The stability provided by the joint commitment is meant to prevent exactly this kind of strategic reasoning: in particular, the obliteration of individual preferences entails that once that I have adopted the collective perspective, and computed the collective solution, I cannot simply turn back to the individualistic point of view and adopt the assumption of your cooperation.

3. Fair defection

One consequence of the holistic character of joint commitment is that the members of the plural subject feel that none of them can rescind the commitment unilaterally, by simply changing her mind. In detail, Gilbert argues that if one has not been given the permission to defect, she, *being rational*, will not defect, because her motivations for action are not due to her personal preferences but rather dependent on the plural subject's goal (Gilbert 2000, pp. 24-25). I will call this process of asking and giving the permission to abandon a joint commitment "fair defection". Fair defection is meant to be the proper way to abandon a joint commitment, in contrast with simple defection, which allows the "abandoned" members to rebuke the defector. Assuming the descriptive adequacy of Gilbert's claim, I will now examine the rationality of fair defection. For the sake of simplicity, I will refer to the PD illustrated in the previous section, though I think that the conclusions (with minor modifications) are valid for all the cases in which there is a joint commitment.
The description of joint commitment in terms of agency transformation suggests that when we consider specific features of collective actions we should distinguish among individual rationality and collective rationality.
Consider the two agents P1 and P2, facing a PD. Imagine that they are jointly committed to implement the collectively efficient outcome (C,C). At a certain time, P1 undertakes the procedure for fair defection, by asking P2 the permission to abandon the joint commitment. Let me focus first on P1 asking P2 the permission to defect, and look at the situation from the individualistic perspective. The first problem is that the adoption of joint commitment prevents P1 from referring to personal preferences as reasons to abandon the collective point of view: if P1's individual preferences have been obliterated, she has no reason to desire defection. For the sake of the argument, let me still assume that P1 can for some reason turn back to individual preferences. As I observed above, is rational for P1 to defect if she *knows* that P2 adopts collective rationality (and thus, plays C); in order to rely on P2 doing her part in the collective solution, however, P1 should better not communicate her decision to stop doing her part in the collective solution.
Also, from the point of view of collective rationality, there are no reasons why a member should prefer defection: the collective solution is the best for the plural subject, regardless the distribution of the gain.
Thus, neither collective nor individual rationality command asking the permission for fair defection: collective rationality prescribes to avoid defection at all, individual rationality prescribes defection without communication for strategic reasons.

Now, turn to P2 reasons to give P1 the permission to defect. If we consider P2 reasons from individual rationality perspective, there is a problem of regression: if P2 adopts individual rationality in deciding whether to give P1 the permission to defect, then P2 has in turn already abandoned collective rationality; if P2 asked P1 the permission to defect, there is a regression problem, while if P2 simply abandoned the joint commitment, then P1 does not need to ask the permission to defect (since the joint commitment is already broken). However, we might admit that P2 can adopt individual rationality, for example, because P1's request signals that the joint commitment has lost the required stability. In this case, individual rationality commands P2 (i) not to give P1 the permission to defect and (ii) to defect: P2 should forbid P1 to defect, so she can rely on P1 cooperation; in this way, P2 can do D and obtain her preferred outcome (C,D).

The other option is that P2 adopts collective rationality in considering whether to give P1 the permission to defect. Yet, from the collective point of view, giving P1 the permission to defect is never rational. As seen in the previous section, cooperation leads to the collective solution of the game (the collectively efficient profile of action), so the plural subject cannot maximize its utility if one of the members ceases to act in accord with the joint commitment. Thus, from the point of view of collective rationality, P2 should not give P1 the permission to defect. As a conclusion, neither individual nor collective rationality allow giving the permission to defect.

The above considerations hold only to the extent that personal preferences are obliterated and collective notions are interpreted in a non-summative and non-correlative way. The conclusion is that from the point of view of individual rationality, defection is not only permitted, but required, though it cannot be properly be considered an instance of fair defection in the sense illustrated by Gilbert. From the point of view of collective rationality, defection on the part of one member is never rational – and each member should do everything in her power to avoid the others' defections.

4. The inclinations plus joint commitment account

So far, I have emphasized that the obliteration of individual preferences is a crucial element of Gilbert's account of joint commitment. Also, I pointed out that this account of joint commitment is inconsistent with the rationality of fair defection.

In her recent book, *Joint Commitment. How We Make the Social World* (2013), Gilbert restates her thesis that joint commitment provides stability to collective actions, by preempting a decision contrary to the collective interest (Gilbert 2013, p. 93).

With specific reference to PD, Gilbert specifies that the adoption of the collective point of view entails that the players accept to do their part in a combination of actions that do not give them what they are most inclined to get. In general, one of the points that Gilbert makes about the motivational force of joint commitment concerns its capacity to "lead to *relatively good outcomes* for all in collective action problems of all kinds" (Gilbert 2013, p. 93; my emphasis). The main problem with this claim is that getting a "relatively good outcome" is not the goal of rationality: by definition, a rational agent does not look at the "relatively good outcomes for all", nor to a "relatively good outcome for herself", but rather at the maximization of her own utility. Note that this holds not only for the individual agents, but also for the collective agent: the plural subject utility (if we persist in a Gilbertean non-summative approach) is insensitive to the distribution of utility among the members; the plural subject, thus, is not interested in the collective solution being "relatively good for all", but rather in its being the best solution for the whole – the collective agent. Despite the problems raised by this specific formulation, however, the point is perfectly consistent with Gilbert's classical account.

I want now to focus on another element, which Gilbert – quite surprisingly – adds to her account in the new formulation. I refer to the introduction of personal inclinations in the explanation of the actions of the parties in a plural subject. As I showed in the previous

sections, the obliteration of individual preferences poses severe limits on Gilbert's account. As a matter of fact, each participant in the activity has personal goals and preferences that differ, even substantially, both from the collective's and from the other participants' goals and preferences.
Perhaps, Gilbert has precisely the intent to avoid these consequences when she says that subjection to a joint commitment does not prevent one's acting according to *her own* best judgment, and that joint commitment does not obliterate one's inclinations. She goes on advancing the proposal of an "inclinations-plus-joint-commitment model of action", on the assumption that such a model might "explain how, though *rationality* requires one to act in a particular way, there may remain a pull in the direction of acting contrary to reason's dictate" (Gilbert 2013, p. 93).
Gilbert tries to combine two elements: the first one is the non-summative notion of joint commitment and the non-correlative apparatus that characterize her account; the other one is the respect of individual rationality.

> Participation in collective agency – in particular, subjection to the underlying joint commitment – does not leave me free to do as I please, *from a rational point of view*. Among other things, it gives me *sufficient reasons to act in a certain way*, reasons I cannot remove at will. Second, it does not – how could it? – deprive me of my capacity to reason and to act according to my own best judgment. I may break away from a collective action in progress at any time – sometimes this may be rationally required, sometimes at least rationally permitted, sometimes not (Gilbert 2013, p. 91).

In the above quotation, Gilbert makes large use of the notion of "rationality". It is unclear whether she is referring to individual or collective rationality, but the meaning of her claims changes drastically depending on the interpretation that we adopt. For example, she says that the subjection to a joint commitment doe not leave me free to do as I please *from a rational point of view*. If I am adopting collective rationality, this claim is trivially true, while if I am adopting individual rationality it is false, for the reasons investigated in the previous sections.
Also, Gilbert argues that one's breaking away from a collective action is sometimes rationally required (or permitted). Yet, the analysis of fair defection provided in *section 3*, showed that individual rationality does always command defection, while collective rationality always preempt defection.
Does the introduction of personal inclinations introduce substantial changes in this analysis? It seems that the possibility of such a change depends on the relations among personal inclinations and joint commitment.
Though Gilbert has merely sketched the inclination-plus-joint-commitment account, without giving much details (indeed, she does not say many thing besides those in the above quotations), I think that with respect to the relation among inclinations and joint commitment there are two main options. The first option is that individual agents involved in the collective activity do not adopt the collective point of view, but rather experiment a change in their individual payoffs due to their perceived relations with the other members. Yet, this explanation is inconsistent with Gilbert's approach, since it removes the role of joint commitment and reduces the problem to one to be solved by individual rationality. The second option holds that individual agents do in fact experiment agency transformation, but – contrary to the holistic interpretation outlined above – the collective payoff is *sensitive* to the distribution of the utility among the players: not only in the sense that a great deal of inequality might prevent the formation of a joint commitment, but also in the sense that the collective payoffs is not a monolithic value, blind to individual gains. This explanation seems

plausible, for it saves both joint commitment and a distinctive role for collective rationality. Yet, it is inconsistent with Gilbert's account of joint commitment, because such a solution rejects non-summativism and non-correlativism.
In conclusion, it seems that there is not a straightforward way to introduce inclinations in Gilbert's explanation of the motivational role of joint commitment, without bringing inconsistencies within the original view.

5. Conclusion

I showed that one problem with Gilbert's classical account of joint commitment is that it conceptually requires the disappearance of individual preferences. I argued that the obliteration of individual preferences makes unintelligible the phenomenon of fair defection, which is the only way to exit a joint commitment. Then, I argued that the re-introduction of individual preferences recently sketched by Gilbert (2013) cannot respect the holistic spirit of her account. The introduction of an inclinations plus joint commitment account is hardly consistent with Gilbert's general theory. In particular, there are two main risks, for one might exaggerate with the import of joint commitment to the detriment of individual preferences (and rationality), turning back to the classical joint commitment account and its limits; or, on the other hand, the introduction of individual preferences might result in the erasure of the stabilizing role of joint commitment.
I think that the problem examined here is part of a general and unsolved problem for the non-correlative accounts of sociality, concerning the relation among the individual and the collective level of explanation. Arguably, the problem of collective rationality as presented in this paper is likely to find a solution only after a general clarification of such relations.

REFERENCES

Bacharach, M. (2006), *Beyond Individual Choice: Teams and Frames in Game Theory*, N. Gold and R. Sugden (eds.), Princeton University Press, Princeton;
Gilbert, M. (1989), *On Social Facts*, Routledge, London;
----. (1996), *Living Together: Rationality, Sociality, and Obligation*, Rowman and Littlefield, Lanham, MD;
----. (2000), *Sociality and Responsibility: New Essays in Plural Subject Theory*, Rowman and Littlefield, Lanham, MD;
----. (2006), *A Theory of Political Obligation: Membership, Commitment, and the Bonds of Society*, Oxford University Press, Oxford;
----. (2013) *Joint Commitment. How We Make the Social World*, Oxford University Press, Oxford;
List, Ch. & Pettit, Ph. (2011), *Group Agency*, Oxford University Press, Oxford.